DOCUMENT RESUME

ED 322 195 TM 015 386

TITLE	Assessment of Student Performance Based on Utah's Core Curriculum: A Handbook.
INSTITUTION	Utah State Office of Education, Salt Lake City.
PUB DATE	Mar 88
NOTE	51p.; Prepared by the Core Curriculum Assessment Planning Committee.
PUB TYPE	Guides - Non-Classroom Use (055)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Academic Achievement; Achievement Tests; *Core Curriculum; Elementary Secondary Education; Graduation Requirements; Performance; Standardized Tests; *State Programs; State Standards; *Testing Programs
IDENTIFIERS	*Curriculum Based Assessment; Utah

ABSTRACT

This handbook is designed to help Utah educators implement student assessment procedures that are in harmony with the State graduation requirements of 1984, which identify implementation and assessment of student achievement as the responsibilities of local boards of education. The State graduation requirements and programs of study allow considerable flexibility for school districts and teachers in choosing assessment techniques to measure student performance in relation to the core curriculum. This handbook provides a set of strategies for educators and a conceptual framework to assist administrators and teachers in implementing the assessment provisions of the elementary and secondary programs of study and graduation requirements. Topics include: critical components for assessment of student progress in relation to core curriculum objectives, components necessary for measuring individual student performance in relation to core curriculum objectives, approaches for measuring student performance based on the core curriculum, teacher-made tests, textbook tests, teacher evaluation of student products and performance, district-constructed tests and item pools, criterion-referenced tests produced by publishing companies and research and development organizations, item pools and end-of-level/course tests developed by the Utah State Office of Education, norm-referenced standardized tests, identification of an instructional management system, and the issue of mastery determination. A glossary, characteristics of tests based on testing purpose, and strategies of assessment for Utah's core curriculum (in the areas of assessment of art, healthy life-styles, information technology, language arts, mathematics, music, science, and social studies) are appended. (RLC)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ASSESSMENT OF STUDENT PERFORMANCE BASED ON UTAH'S CORE CURRICULUM

- A HANDBOOK -

THE UTAH STATE OFFICE OF EDUCATION

James R. Moss
State Superintendent of Public Instruction

Kolene F. Granger
Associate Superintendent, Research and Development

Prepared By:
The Core Curriculum Assessment Planning Committee

David E. Nelson, Chairman

LaMar Allred
Gary Carlston
Jay Donaldson
Briant Farnsworth
Jim Henderson
Lynn Jensen
Max Jensen
Nancy Mathews
Bonnie Morgan
John Ross
Devon Sanderson
Harold Shaw

March, 1988

UTAH STATE BOARD OF EDUCATION

250 East 500 South
Salt Lake City, Utah 84111

Keith R. Checketts, Chairman
Logan, Utah 84321
District 2

John M. R. Covey
Bountiful, Utah 84010
District 3

Ruth Hardy Funk, V[illegible]hairman
Salt Lake City, Ut[illegible] 84109
District 4

Darlene C. Hutchison
Salt Lake City, Utah 84119
District 5

Neola Brown
Beaver, Utah 84713
District 1

Valerie J. Kelson
Midway, Utah 84049
District 9

Donald G. Christensen
Salt Lake City, Utah 84120
District 7

M. Richard Maxfield
Salt Lake City, Utah 84117
District 6

Margaret R. Nelson
Provo, Utah 84603
District 8

James R. Moss, Executive Officer

Twila Bringhurst, Secretary

FOREWORD

This handbook, Assessment of Student Performance Based on Utah's Core Curriculum, has been prepared to help Utah educators implement student assessment procedures which are in harmony with the state graduation requirements of 1984.

The state graduation requirements and programs of study allow considerable flexibility for school districts and teachers in choosing assessment techniques to measure student performance in relationship to the core curriculum. This handbook provides a set of strategies which will be helpful to educators in accomplishing student assessment tasks. Also provided is a conceptual framework to assist administrators and teachers in implementing the assessment provisions.

This handbook is one part of an ambitious program of measurement development undertaken by the State Office of Education in partnership with Utah school districts. It is our hope that products like this one, in conjunction with test item pools and criterion-referenced tests which are being developed and provided to Utah educators, will be a sound basis for building effective student assessment systems at both the classroom and district levels.

This handbook was produced by the Core Curriculum Assessment Planning Committee, chaired by Dr. David E. Nelson.

James R. Moss
State Superintendent of Public Instruction

TABLE OF CONTENTS

	Page #
Introduction and Background	1
Critical Components for Assessment of Student Progress in Relationship to Core Curriculum Objectives	2
Components Necessary For Measuring Individual Student Performance in Relationship to Core Curriculum Objectives	4
Approaches For Measuring Student Performance Based on the Core Curriculum	5
Teacher-Made Tests	5
Textbook Tests	5
Teacher Evaluation of Student Products and Performances	5
District-Constructed Tests and Item Pools	6
Criterion-Referenced Tests Produced by Publishing Companies and Research and Development Organizations	7
Item Pools and End-of-Level/Course Tests Developed by USOE	7
Norm-Referenced Standardized Tests	9
Identification of an Instructional Management System	9
The Issue of Mastery Determination	11
Appendix A - Glossary	15
Appendix B - Characteristics of Tests Based on Testing Purpose	23
Appendix C - Strategies of Assessment For Utah's Core Curriculum	33

Introduction and Background

In January of 1984 the State Board of Education passed a new set of graduation requirements for Utah's public schools along with a set of requirements for elementary and secondary school programs of studies. This policy departed from previous graduation requirements by making specific statements about standards for both elementary and secondary education and establishing an objectives-based core curriculum to be prepared by the State Office of Education with input and assistance from local school districts.

Implementation and assessment of student achievement in respect to the core curriculum are identified as the responsibilities of local boards of education. The state requirements specify that "informal assessment" should be ongoing and that "formal assessment" shall occur during or at the completion of the primary grades (K-3), the intermediate grades (4-6), and grades 8, 10, and 12 at the secondary level.

This paper provides a conceptual framework to assist Utah school districts in implementing the assessment provisions of the elementary and secondary programs of study and graduation requirements. It is one product in a series of efforts undertaken by the State Office of Education to clarify and interpret assessment issues, especially as they pertain to the new requirements and the core curriculum. Previous reports have examined current testing and assessment programs within the state and across the nation and analyzed possible statewide thrusts in the testing and assessment area. Yet another effort, related to this matter, was a study of the reactions of Utah school districts to a variety of assessment options identified in the papers just mentioned and the collection of information pertaining to the perceived

needs of Utah school districts relative to assessment of the core curriculum and other testing activities.

The present strategy of the State Office of Education, as called for by the State Board of Education, is to provide additional leadership and support to local school districts in strengthing their own assessment efforts. The State Office of Education has undertaken an ambitious program to develop item pools and end-of-level and end-of-course tests which are specifically based on the core curriculum. The focus of this paper is specifically on providing additional definition and clarification of the core curriculum assessment requirements to assist districts in implementing their own programs of assessment related to the core.

Critical Components for Assessment of Student Progress in Relationship to Core Curriculum Objectives

The implementation of the core curriculum is obviously a major effort which impacts on every level of Utah's public educational system and involves numerous elements of the system. At the very least, implementation involves educators becoming familiar with the basic core curriculum document, analyzing curriculum and instructional approaches to ensure core objectives are included, and identifying appropriate measurement techniques to assess student performance in relationship to core objectives.

Several aspects of the overall implementation of the core are particularly crucial, both from the standpoint of instruction as well as assessing student progress. These aspects include a set of measurement approaches to gauge performance on objectives and an instructional management system which includes core standards and objectives. The following illustration, "A Model for Core Assessment," shows the important components necessary for a systematic approach to assessment based on the core curriculum.

A MODEL FOR CORE ASSESSMENT

IMPORTANT IMPLEMENTATION CONSIDERATIONS

ASSESSMENT COMPONENTS	1. FREQUENCY OF USE OF COMPONENT	2. WHO MANAGES COMPONENT	3. SUPPORT MECHANISMS WHICH CAN BE USED TO IMPLEMENT COMPONENT
I. PRE-ASSESSMENT/ DIAGNOSIS/ PLACEMENT	Beginning of Year or Course and as Needed Throughout Year	Classroom Teacher With Support From School and District	(1) Standardized and Locally Developed Diagnostic/Placement Tests (2) Teacher Observation (3) Results From Ongoing Assessments (Progress Monitoring and Mastery Determination)
II. PROGRESS MONITORING/ UNIT MASTERY TESTING	Continuous Throughout The Year or Course	Classroom Teacher With Support From School and District	(1) Teacher-Made Tests (2) Textbook Tests (3) USOE Core Assessment Item Pools (4) Teacher Evaluation of Student Products and Performances
III. FORMAL ASSESSMENT	After a Substantial Portion of a Course is Completed in Most Instances	District Assessment System For Specific Areas/ Objectives --- Classroom Teacher for Specific Areas/Objectives	(1) Criterion-Referenced Tests and Performance Evaluations. A. Commercially Developed B. Locally Developed C. USOE Developed End-of-Level/Course Tests
IV. INSTRUCTIONAL MANAGEMENT SYSTEM/ INDIVIDUAL STUDENT RECORD KEEPING	Continuous Throughout the Year or Course	Classroom Teacher with Support From School and District	(1) Manual or Computer Managed Instruction System to Document Student Progress and Performance

Components Necessary For Measuring Individual Student Performance in Relationship to Core Curriculum Objectives

The state graduation requirements and programs of study allow considerable flexibility for individual school districts in choosing assessment techniques to measure student performance in relationship to the core curriculum. The following sections describe several different approaches to gauging performance based on core objectives. In examining various approaches to student assessment, it may be useful to consider some major assessment components or purposes which are relevant to an objectives-based curriculum. The three major kinds of assessment needed to support instruction based on objectives include the following:

I. Pre-assessment. This includes surveying student knowledge and capabilities prior to instruction for the purpose of identifying individual and total class instructional needs. Another dimension of pre-assessment could entail administering a very comprehensive and formal set of assessment procedures which would allow students to "test out" of a particular course.

II. Progress Monitoring/Unit Mastery Testing. This is a procedure in which teachers check frequently on the progress of individual students or an entire class through approaches such as brief tests, observation of performance, analysis of assignments, and a variety of other means. From an instructional point of view, it is probably most sensible to accomplish unit mastery testing relatively soon after the completion of instruction on a particular objective or group of related objectives such as a unit of instruction. Such an approach ensures that students who have not mastered a particular objective or set of objectives can get additional help in these areas.

III. Formal Assessment. This aspect of assessment is intended to document attainment or non-attainment of specific objectives over a significant period of time or content. This is also the aspect of performance evaluation which needs to be the most technically precise element of the assessment system.

Included in the area of formal assessment would be end-of-term, end-of-year, or end-of-course comprehensive, objectives-referenced tests which would cover the significant content of a particular curriculum area or course for the identified period of time. Indeed, some combination of mastery determination close to the completion of instruction in conjunction with the end-of-term or end-of-course approach would represent a very comprehensive strategy in documenting both short-term and long-term accomplishments of students.

Approaches For Measuring Student Performance Based On The Core Curriculum

The following sections describe some specific approaches to evaluating student performance in relationship to objectives which might be considered in implementing the three major types of assessment discussed above.

Teacher-Made Tests

Teacher-made tests constructed by individuals or by groups of teachers might represent a strong component of a total objectives-based assessment system. Such tests may be aligned very closely to the curriculum of a particular teacher or school in a way that no pre-packaged, commercial test is likely to be. If used for mastery determination, it is important that teacher-made tests be developed carefully so that judgments based on test performance provide a reliable indicator of student status. Careful review of test item content, along with actually trying out such tests in a pilot situation, constitute good practice in creating such instruments.

Textbook Tests

Tests which are included in textbooks may be useful for assessing objective attainment to the degree that they have several items which measures an identifiable objective. A common problem with such tests is their tendency to include only an item or two related to a specific objective. While this does not represent a problem for "informal" assessment purposes such as progress monitoring, it would render such a test of very little use for mastery determination.

Teacher Evaluation of Student Products and Performances

Many objectives in the state core curriculum are not subject to being measured through the use of traditional paper-and-pencil achievement tests. In such areas as speaking, listening, writing, the arts, vocational educa-

tion, and other areas, it is the creation of a product or the demonstration of a set of behaviors which is key to achievement of a particular objective. In these instances, teacher judgment may be the most valid measure of student performances or products. As with any other form of student assessment which is used to certify attainment of an objective or set of objectives, procedures based on teacher judgment should be developed carefully and include specific criteria or descriptors against which a students product or performance may be compared. Such a procedure might take the form of a rating scale in which several levels of performance are identified ranging from failure to achieve the objective to superior performance. To the extent each point on such a scale is defined specifically, the reliability of such judgmental procedures is enhanced.

District-Constructed Tests and Item Pools

Some Utah school districts already have in place systems of objectives-referenced testing which have been constructed specifically to measure performance in respect to the district curriculum. For example, Jordan district with its Goal Based Management System (GEMS) and related tests, as well as a rather comprehensive set of minimal competency tests, has in place numerous student evaluation techniques which would produce information relevant to the state's core objective structure. Recent efforts in other districts including Alpine, Granite, Murray, and Washington, are focused on building measurement systems specifically targeted on district core objectives. Such district-wide systems represent a substantial plus where they exist when tests and items are already keyed to a set of objectives which includes the state core. In addition, tests which have been used over a period of time and have demonstrated acceptable content and technical properties add a degree of rigor to such a system.

Criterion-Referenced Tests Produced by Publishing Companies and Research and Development Organizations

Two different kinds of products are available from these organizations. The first product is an intact test, usually measuring basic skills in areas such as reading and mathematics at the elementary levels. Many of these tests are designed for pre-assessment applications while others may have a primarily end-of-course assessment purpose. Still other total testing packages may cover the range of applications from pre-assessment through mastery determination. A key issue in the use of "off the shelf" criterion-referenced tests is the relationship between the objectives underlying such tests and the core curriculum. While a reasonable degree of relationship may be expected in such areas as mathematics and reading, tests in other areas at the elementary level, or end-of-course tests produced for the secondary level may not be closely aligned with the core curriculum.

The other major type of product available from test publishing firms and some other sources is a custom-built, criterion-referenced testing procedure in which local educators can specify objectives and have tests prepared with items which match those objectives. These types of services are, again, provided mainly in the area of academic basic skills and also tend to be most prevalent in elementary mathematics and reading.

Item Pools and End-of-Level/Course Tests Developed by USOE

The State Office of Education, in conjunction with local school districts, universities, and other research organizations, has an ambitious schedule of item pool and test development projects under way. There are two basic kinds of products being developed by the State Office in conjunction with these other groups. They are the following:

End-of-Level and End-of-Course Tests. The State Office has been working with several school districts and Utah universities as well as the U-SAIL project and the Northwest Regional Educational Laboratory in developing end-of-level and end-of-course tests. These tests are being developed with great technical precision and are being field tested at least two and typically three times. The end-of-level/course tests have two major purposes: First, they provide a final check on student attainment of core content. Second, they will help document program strengths and weaknesses. These are "formal" tests and should be treated in the same manner as other important tests such as norm-referenced test batteries.

At the same time these tests are being developed, the USOE data processing staff is creating a sophisticated test scoring and reporting system, which will greatly facilitate using the tests and will create reports much like those which are currently available from USOE for norm-referenced tests. This system should be ready for use early in the spring of 1988. Districts would, of course, be at liberty to score these tests locally since complete answer keys will be provided for all tests.

Item Pools. The other major type of product being developed in conjunction with districts and universities, as well as other groups, is a set of item pools for use by teachers and districts in more informal, day-to-day or unit-oriented, mastery assessment. The first such product is an item pool for elementary mathematics, grades 1 through 6. Development of other item pools, as well as a direct writing assessment project in the language arts, is well under way and initial drafts of such item pools have been prepared in several areas. The unique aspect of these item pools is that they will be completely coded to the core curriculum. Appendix B of this report presents a set of guidelines which might be used in constructing various kinds of tests from item pools.

Norm-Referenced Standardized Tests

While the major purposes for administering norm-referenced, standardized tests deal with general monitoring of the performance of individual students, buildings, or a total system, such instruments also produce information which may be interpreted in terms of specific objectives. A typical norm-referenced test can provide only limited information about the achievement of specific objectives. The structure of such tests dictates that only a few items may measure any particular objective. Norm-referenced tests do, on the other hand, provide a survey of student knowledge across a rather broad-range of content difficulty in areas such as math, reading, English, science, and social studies. Thus, the maintenance of a quality norm-referenced testing program remains important for any school district in order to track performance across a wide range of skills and over extended periods of time for purposes such as total system monitoring and program evaluation. As a component in measuring individual student performance in respect to the core curriculum, however, the typical norm-referenced test provides, at best, a rather incomplete picture of performance across the total set of objectives in the core.

Identification of an Instructional Management System

A necessity for any objectives-based curriculum is the existence of a record keeping system which allows teachers and students to know precisely what objectives are expected to be achieved and provides a record of progress in achieving those objectives. Such a management or record keeping system can be as simple as a checklist which details the objectives of a grade level or course and allows the teacher to indicate student progress in attaining objectives. Recording progress with a basic approach such as this might

range from entering test scores, to noting summary judgments based on a scale ("superior performance", "good performance", "objective not achieved"), to a checkmark indicating attainment or non-attainment of a given objective. The assumption with the latter procedure is that test scores or other performance evidence exists to substantiate the use of the simple checkmark notation.

This basic paper-and-pencil approach to student record keeping is obviously already familiar to most teachers. The critical difference between a traditional student record keeping system and one which is needed to support the implementation of the core curriculum is the core objective structure and the ability to relate student evaluation specifically to that objective structure.

Beyond the rather straightforward, manual student record keeping approaches described above, several Utah school districts have had significant amounts of experience with computer-managed instruction. In its simplest form, a computer-managed instructional system is simply an automated way of maintaining information describing attainment of objectives by students. Computer-managed instruction (CMI) also opens up numerous opportunities for analyzing information about student objective attainment which is somewhat laborious using a manual, paper-and-pencil record keeping system. For example, many CMI systems will quickly identify instructional groups by listing students who have not obtained a particular objective or set of objectives. CMI systems also may incorporate test scoring capability, test item generation, and a variety of ways of profiling the performance of individuals and groups of students. As with the manual record keeping system, a computer-managed system must be loaded with core objectives which may then be assessed in any number of ways from formal testing through teacher judgment.

The Issue of Mastery Determination

Perhaps no other aspect of objectives-based instruction and measurement is so slippery or controversial as the matter of determining student attainment of objectives or mastery. Many professionals, in both psychometrics and education, argue that setting mastery levels or performance standards can never be accomplished in a totally technically or conceptually defensible manner.

Lorrie Shephard (1984), in an excellent analysis of standard setting and related issues, offers the following practical advice to educators who must, in spite of technical and conceptual problems, set mastery levels:

> Faced with the problem of setting passing scores, teachers might ask these questions about the test: Is the test highly homogeneous? (That is, are the items all parallel versions of one another? For example, are they all one digit multiplications by nines?) Is the purpose for testing to decide on the next day's instruction or to assign grades? Only when the tests are highly homogeneous do insights derive from state mastery models. With math mastery tests, for example, it is possible to imagine that a child has either mastered his nine's table or not. (Note that partial knowledge is simply defined as non-mastery.) On highly homogeneous tests, mastery is defined as 100% with some allowance for measurement error. There are no magical or particularly scientific ratios here. The conventional 80% maybe too high or too low. The teacher simply tries to judge how many careless errors to expect typically from pupils who are clearly masters.
>
> For subject areas other than math, content cannot be divided into homogeneous bits. Furthermore, test questions covering identical instructional objectives can be either hard or easy. Therefore, standards should not exist independent of the test. We should be suspicious of testing programs where the standard is a fixed value, always 80% or 70%. When standards are highly arbitrary (set without regard to the test), apparent successes and failures may merely be the result of changes in the strictness of the standard. [pg. 190]

A major conclusion which may be drawn from Shephard's analysis of setting mastery levels is that such standards should always be subject to review based on the actual performance of students taking a particular test. It is certainly conceivable that for a very fundamental objective measured by rather easy items that a standard of 90% could actually be too low to identify mastery. Conversely, for another objective measured by relatively difficult items, a standard of 60% may be all but unattainable for a great majority of students.

In the development of lengthy, formal objectives-based tests used to make decisions about grade advancement or graduation, Shephard suggests a comprehensive composite model for setting standards which would include the following elements:

(1) absolute judgments based on inspection of test questions,
(2) empirical validation data based on judged masters and non masters,
(3) decisions about acceptable passing rates, and
(4) adjustments for unreliability to minimize the costs of classification errors.

It is almost always good practice when tests are used to make crucial decisions that standards or passing scores be set based on the total test rather than its components. The reliability of judgments based on large numbers of items is nearly always greater than judgments based on smaller numbers of items.

REFERENCE

Shephard, L. Setting performance standards. In Berk, R.A. (ed), A Guide to Criterion-Referenced Test Construction. Baltimore: Johns Hopkins University Press, 1984.

APPENDIX A

APPENDIX A

GLOSSARY

Assessment: Assessment, as the term is typically used in education, means the process of gathering meaningful information about some important student characteristic. Student characteristics include, but are not limited to, academic achievement, attitudes, aptitudes, and personality characteristics. The term assessment is very close in meaning to the term measurement, which is commonly defined as the process of obtaining a numerical description of the extent to which a person or thing possesses some characteristic. Although both the terms assessment and measurement are often equated with paper-and-pencil testing, this is a very narrow conceptualization of the process which could also include such techniques as actual performance evaluation, observational assessment, and oral assessment, to name only a few.

Informal Assessment: As this term is used in Utah's 1984 graduation requirements and elementary and secondary programs of study, it means the gathering of information about student academic performance and products on a regular basis so that both teachers and students have a clear picture of student progress in relationship to specific instructional objectives. "Progress Monitoring" and "Unit Mastery Testing" are both part of "Informal Assessment." Key characteristics of informal assessment are that it is frequent, provides information to make instructional decisions by the teacher, and provides feedback to students. As with the general term, "assessment", informal assessment may be accomplished through a variety of means ranging from brief tests through teacher questioning or observation of student performance.

Formal Assessment: Formal assessment is the process of gathering cumulative information on student performance in order to certify the attainment of a particular objective or set of objectives. As this term is used in the 1984 graduation requirements, it implies a set of measurement procedures which certify student mastery of the core curriculum during or at the completion of grades 3, 6, 8, 10, and 12.[1] Such assessment, used for purposes of mastery determination, would be administered under uniform conditions and employ valid and reliable measurement techniques to provide information for making judgments about the performance of individual students. As with other forms of assessment, formal assessment should certainly not be limited to the typical paper-and-pencil test and could conceivably be based on all of the measurement possibilities already discussed in the general definition of "assessment" presented above. Formal assessment would typically be more general than informal assessment and would be used primarily to evaluate the degree to which substantial numbers of objectives had been mastered over an entire year or course or some other substantial portion of time. Because of the long-range focus of formal assessments, they might sample only the most critical major outcomes in a particular area rather than all the specific details of finite objectives.

---

[1] Operationally, formal assessment at the secondary level might best be approached through end-of-course tests rather than by some blanket grade level test which could not be sensitive to different programs of study.

Cut-Off Scores/Mastery Determination: When determining mastery, a cut-off score is a numerical standard which a student must obtain in order to demonstrate mastery of a particular objective or set of objectives. The concept is a quantitative one and is one most easily applied to assessment procedures which produce clear quantitative results such as paper-and-pencil tests. Setting of a cut-off or mastery score usually connotes creating two groups; those who have mastered a particular body of information and those who have not. Such a dichotomy is obviously a distortion of the actual state of affairs which finds students ranging in performance from very little understanding of a particular objective to very great understanding. Thus, great care should be exercised in setting mastery levels. Arbitrary mastery levels such as 80% may be too high or too low depending on the nature of a particular test. Some informal and formal procedure have been developed to aid in establishing such scores.

Core Curriculum: The core curriculum consists of the foundation skills, knowledge, and behaviors which, when mastered, provide each student access to a diversity of educational or vocational choices. The core curriculum is not the total curriculum for a given level or course. The total curriculum extends beyond the core and is determined by the district so long as it incorporates the core standards identified.

Core Curriculum Standard: A statement describing a central learning concept of the core curriculum.

Core Curriculum Objectives: Measurable or observable outcomes of learning experiences which relate to the central concept described in a core standard.

Instructional Management System: An instructional management system is any means of consistently tracking what any student currently can or cannot do or what that student does or does not know in relation to a prescribed curriculum. The system could also connect available instructional materials (books, films, etc.) to where the student is (e.g., the teacher knows if the student is working on Standard 6 in Level 6 language arts, which instructional materials directly apply). While many people think "management system" is synonymous with "computerized management system", this is not necessarily the case. Many texts (e.g., basal readers) contain paper-and-pencil management systems which are as effective as current computerized systems.

Mastery: Mastery means that a student has learned an element in the curriculum. A student who has "mastered" an objective or standard can apply that objective or standard (either in school or in the "real" world) or is able to move on to the next objective or standard in a series.

Instruction For Mastery: This is any instructional methodology that focuses on teaching concepts, skills, and attitudes as described in the core curriculum standards. The objectives underlying each standard in the core establish performance levels that the student must display or possess to show that mastery has been achieved. Mastery is based on the philosophical premise that all children can learn. Instruction should take into consideration different learning rates, prerequisite concepts and skills, and clear expectations for student performance. "Mastery Learning" is one approach among many which can be employed to teach to core standards and objectives.

Competency: Competency is the knowledge and skills required to accomplish a task, performance, or assignment at a specified level (cut-off point). The terms competency and mastery are often used synonymously.

Norm-Referenced Tests: Norm-referenced tests provide general monitoring of the performance of either a student or a system. Norm-referenced testing permits comparisons between the performance of a particular student or system with that of a national norm group. For instance, we could compare Jane's reading score with that of other third-grade students across the country. Norm-referenced tests are generally very carefully developed and published by national testing firms.

Criterion-Referenced Tests: A criterion-referenced test determines specifically what students do and do not know in reference to specific educational objectives -- that is, those words Jane can decode and those that she cannot. Such tests (also called objectives-referenced tests) are appropriate for determining where students need additional instruction, placement of students in specific curricular materials, and mastery determination. In general, criterion-referenced tests are created for very specific purposes and, usually, national comparison information is not available for such tests. Criterion-referenced tests can be constructed by teachers, districts, test publishers, etc.

Performance Evaluations: As the name implies, performance evaluations are assessments of what students can and cannot do in reference to specific educational objectives. For example, the specific educational objective might be "the students will make valid inferences from their American history text's

discussion of the cause of the Civil War". Probably the most direct way to assess this objective is to ask the students to demonstrate in speech or writing what they believe caused the American Civil War. The students' responses would then be judged valid or invalid in reference to predetermined criteria and/or previous examples of student responses. Yet another example would be a procedure for assessing a product created in a vocational education class. Logically, there are few national norms for performance evaluations, although they may provide some of the best means for assessing higher-level thinking skills. Teachers, districts, and other informed professionals can construct performance evaluations.

APPENDIX B

A P P E N D I X B

CHARACTERISTICS OF TESTS BASED ON TESTING PURPOSE*

When designing a test or a testing program, it is important to identify the purpose(s) for testing, the evaluation decisions to be made, and who will make the decisions. Several important testing purposes which apply to academic achievement testing are discussed on the following pages. It may, however, be helpful to first review the section on measurement concepts below.

Measurement concepts

There are several key concepts that need to be examined as one considers testing purposes. These include: (1) width of the test, (2) curriculum relevance, (3) passing score precision, (4) test accuracy, and (5) test length.

Width

One important consideration in the design of a test is the range of difficulty or width of the test. If the expected achievement of students represents a wide range and the results of the test are to adequately reflect this range, then an appropriate test would also have a wide range of difficulty in the items. On the other hand, if the achievement of the students is expected to be narrow, or if the decision to be made with the data focuses on a specific level of achievement (e.g., mastery test, placement in talented and gifted program), then the test should have a narrow range.

---

* This material copyright © 1987 Northwest Evaluation Association Inc. Portland Oregon. This section is reproduced with the permission of the Northwest Evaluation Association.

Curriculum Reference

One of the key elements of test validity is whether the content of the test is appropriate to the purpose. In some cases, the content need not be tied closely to the content of a specific instructional program, particularly if the purpose is for placement. However, in other cases it is extremely important that the content of the test match a specific curriculum, especially for mastery and competency tests.

Passing Score Precision

When a single test is to be used to make decisions about whether students are above or below a cut-off or passing score, it is extremely important that the test give a precise score in the range around the cut-off score. This test would have a predominate number of test items that cluster around the difficulty level of the cut-off scores with few items above and below the cut-off level. Tests with high precision tend to have a narrow range. These tests often do not give an accurate score for students whose "true score" is well above or below the cut-off score, but the test is precise in determining whether a student is above or below the cut-off.

Test Accuracy or Acceptable Error

Whereas precision refers to how well a test discriminates between students scoring above and below a cut-off, accuracy refers to how well the test gives an accurate score for a student regardless of the student's achievement level. An accurate test is one that has adequate test items that reflect the level of achievement of the student taking the test. A student who answers almost all items on a test either correctly or incorrectly will have a less accurate score than one who scores more at the mid-range of the test.

Basically, the most information is obtained from items close to the achievement level of the student. Items that are extremely easy for a student provide very little real information because he or she will answer most items correctly, except for careless errors. Similarly, items that are extremely difficult for a particular student provide very little real information because he or she will answer most of the items incorrectly, except for chance guessing. Items that are close to the student's level of functioning provide the best information.

To select items close to the achievement level of a student, one must know his or her approximate functioning level before the measurement process begins. Access to this information may be available from past measurement, locator tests, grade level, or other data that would enable one to "tailor" the testing program to an individual student. Focused or tailored testing is a more efficient and humane use of class time.

Length

After the characteristics of a test have been determined, quite often the number of items on the test (or its length) is also determined. For example, a unit test that does not require a high level of accuracy would be a short test. On the other hand, a grade level test that might cover five major goals would need to be longer. The length of the test is determined by a combination of the breadth of curriculum to be assessed, desired accuracy of the test, and width of difficulty of the items on the test. The longest test would be one that covers a number of goals for students with a wide-range of ability and would require a high level of accuracy.

Testing Purposes

Unit Tests

Most course or grade level instruction is broken down into a series of units that cover one to several goals. Unit tests are often constructed to determine how well students have learned the content of the unit. Typically, teachers assign student grades and determine if review of the unit is necessary based on the results of a unit test. Unit tests can be relatively short (15-30 items). The items have a relatively narrow width of difficulty and relate directly to the content of the curriculum. However, unit tests are often not as accurate as longer tests.

Mastery Tests

Mastery learning programs typically include curriculum imbedded tests which are used to determine whether a student has met or surpassed (mastered) a specified criterion and will thus be allowed to proceed to the next unit of learning. These tests are specifically related to the curricula, are usually rather brief and include pass/no pass criteria. As with competency tests, mastery tests focus on a very narrow range of measurement and are developed with multiple parallel forms; however, they do not require the same level of precision.

Course/Grade Level Tests

It is difficult to find commercially developed tests that adequately assess the curriculum in a course or grade level, because they often do not match local curricula. Tests for grade level or course curricula are usually developed from established item banks or are written and refined locally. In addition to assessing the effectiveness of curricula, these tests can be used

for individual performance monitoring and provide teachers with information which can be helpful when establishing mid-term or end-of-term grades. These tests require a direct relationship to local curricula, are designed with a narrow range of measurement, can provide for good accuracy at the test score and goal level and can provide scale scores along with local or classroom norm data.

Survey or Functional Level Tests

In curriculum areas such as reading, mathematics and writing, student achievement can be viewed as progressing along a continuum where students learn skills that build upon other skills, or where students learn to perform more complex versions of the same task. Tests to monitor student growth can be used to determine if a student's progress during a given year is comparable to past years. Uneven patterns of growth may indicate that a student is experiencing difficulties with the school experience or perhaps difficulties outside the school.

However, most students do not learn at a constant rate even when they are not experiencing unusual stress. Learning seems to be a series of rapid growth periods and periods of stabilization, with little or no growth in a particular subject area. The latter may be a time of consolidation of previously learned material or it may be that a student's mental resources are being diverted to other subject areas. Whatever the reason, if these growth spurts can be detected, it may be possible to adjust the learning environment to better achieve the desired educational goals.

Tests for monitoring student growth usually come in two forms. The first type, survey tests, have a broad range of measurement with high accuracy at the total test score level but low accuracy at an individual goal level. These tests are commonly based on local and/or publisher norms.

While population (norm) scores are most commonly used with these tests, scale scores are also available and assist with monitoring growth from year to year.

The other form of test used to monitor student growth is called a functional level test. Unlike the survey test, the functional level test assesses a narrower range of achievement but is geared more to the functioning level of the student. Whereas a survey test may have one test that covers more than one grade level, a functional level test system will have a series of three or more overlapping tests. The student takes the one form that best approximates his or her level. The scores from these tests are commonly on a scale that can be used to determine growth from year to year. This scale can also be referenced to a norm group for determining population scores. These tests can provide high accuracy at the total score and goal level if the test is developed to match local curriculum.

Screening and Placement Tests

It is often necessary to identify students who are to participate in various special programs (e.g., special education, Chapter 1 or talented and gifted programs). In other cases, students are selected for courses or instructional groups based on achievement levels within the relevant curriculum. In either case, it is important that the selection be systematic, uniform and fair. Often students are screened first before they actually take a test for placement. Screening tests give a rough indication of those students who might be in a special program from those who should not be in the program. Screening tests usually have a wide range of difficulty and are less precise than a placement test. A key element of a placement test is its precision at the cut-off point being used to select students, particularly if the test score is the only factor being considered, then the test will need a

broader range of difficulty. Often the content being assessed is not tied closely to a specific instructional program and may not give an accurate score for students near the extreme range of achievement.

Minimum Competency Tests

Minimum competency tests are often used to make decisions about whether a student will be promoted or graduated and frequently includes a remediation and retesting option for those not meeting the established criteria. Generally, these tests are closely related to specified curriculum, developed around a narrow measurement range, include multiple parallel forms and have specified pass/no pass criteria. The critical nature of the judgment to be made with these tests dictates a high degree of precision, i.e., a reasonable number of items focused around the criteria contributing to confidence in an individual's score close to the criteria. Student scores well below or above the pass criteria are usually not very accurate.

The table on the following page illustrates the characteristics of tests designed for specific purposes.

DESIGNING TESTS FOR A PURPOSE

Type of Test	Length	Width	Acceptable Error	Passing Score	Reporting
1. Unit	15-30	Narrow	Large	Maybe	Total or Goal*
2. Mastery	20-30	Narrow	Moderate to Large	Yes	Total or Goal
3. Course or Grade Level	40-60	One Grade	Moderate to Small	Maybe	Goal
4. Survey	40-60	Two-three grades	Moderate	No	Total
5. Functional Level	40-50	Narrow	Small	No	Total and Goal
6. Screening	40-50	Broad	Moderate	Yes	Total
7. Placement	40-50	(Varies)	Small (Varies)	Multiple	Total
8. Minimum Competency	40-50	Narrow	Very Small	Yes	Total

* Tests with goal reporting usually have 6-10 items per goal.

APPENDIX C

A P P E N D I X C

STRATEGIES OF ASSESSMENT FOR UTAH'S CORE CURRICULUM SUBJECT AREAS

ART ASSESSMENT

Assessing student performance in the core area of art is unquestionably a challenging task. Art does not easily lend itself to traditional measurement techniques generally used in the classroom; this subject area will require the extensive use of visual references in order for adequate assessment to occur. Within the scope of Utah's new Core Curriculum the focus of this subject has evolved from a heavily performance-oriented curriculum to one which also places considerable emphasis upon the development of aesthetic understandings and appreciation for art. Students should be provided with comprehensive background so that they can function well in the roles of participant, observer, and critic. A participant role involves dealing actively and directly with the medium of an art form such as painting, drawing, sculpting or sketching. An observer role involves actively responding to the aesthetic qualities of an art object. The role of critic involves responding to the historical significance of the work or its relationship to the culture that produced it. All three roles contribute significantly to the student's development of the knowledge and skills they need to become intelligent, discriminating patrons of the visual arts.

The philosophy of assessment for this subject is that every effort must be made to evaluate a student's artistic knowledge and skills through direct observation and the actual creation of art objects, paper-and-pencil measurements, and by any other means which will yield valid and reliable assessment

information. Such artistic performance endeavors as creating an impressionistic painting, a charcoal sketch, or a piece of sculpture are examples of skills which can be assessed through direct observations and the use of a rating scale. Scoring a student's performance will require the observer (or teacher) to make judgments within the scoring framework of the rating scales. A student's conceptual understandings, knowledge of techniques, perception of aesthetic qualities and knowledge of historical significance (art appreciation) can be assessed, in part, by paper-and-pencil measurements. These measurements will require the knowledgeable input of teachers and other educators expert in the visual arts. The resultant test items will need special question stems and response alternatives specific to various forms of art and will also require the use of numerous illustrations. Assessing this subject area will truly be an art!

HEALTHY LIFESTYLES ASSESSMENT

Healthy lifestyles is one of the many important subject areas included in Utah's Core Curriculum. It encompasses both health education and movement and fitness courses. Health education includes five core strands which address life skills, growing and developing, positive and negative health practices, community health and disease presention, and safety and first aid. Movement and fitness focuses primarily upon the core strands of body/space awareness and management, motor fitness, basic play and sports skills, games and sports, dance forms (creative movement), physical fitness and lifetime activities. Assessment of healthy lifestyles, which must address these various strands, should include measurements of both cognitive and physical learning.

The philosophy of assessment for this subject is that every effort must be made to evaluate a student's healthy lifestyles knowledge and skills through direct observation and physical performance of these skills, paper-and-pencil measurements, and by any other means which will yield valid and reliable assessment information. Such physical tasks as doing push-ups, running a specified distance, walking a balance beam, or lifting weights, are examples of skills which can be assessed through observation coupled with direct measurements such as number of repetitions, elapsed time, and quantities of weights lifted. Other physical tasks such as throwing and catching a ball, performing a gymnastic stunt, doing aerobic exercises, or hitting a backhand tennis shot will likely require structured, well-defined observations by an observer (teacher) to rate such performance. Scoring the student's performance will require observers to make judgments within the framework of scoring scales or specific response options.

The cognitive learning which will occur in the study of healthy lifestyles, including such information as understanding basic health practices, acquiring knowledge of the Food and Drug Act, understanding the harmful effects of alcohol and drugs, knowing the fundamentals of home safety and basic first aid, as well as developing an appreciation and knowledge of how the body works, could be assessed through paper-and-pencil measurements. These measurements will require the knowledgeable input of teachers and other educators skilled in healthy lifestyles, as the various test items will require special question stems and response alternatives specific to health education, physical fitness, and movement.

INFORMATION TECHNOLOGY ASSESSMENT

Assessing student performance in the core area of information technology is a task which will require some innovative measurement. The standards and objectives for information technology may be taught as specific classes or may be infused into other areas of the curriculum, which will require assessment techniques which can be applied in either setting. Information technologies, and specifically computer literacy, enable students to function effectively in an information-based society. Students should develop the skills necessary for good communication, and to effectively use technology to facilitate their learning throughout life. Utah students should also realize the implications and potential of technology relative to an information society and the world in which they live.

The philosophy of assessment for this area of the core curriculum is that a substantial effort must be made to evaluate a student's information technology skills and knowledge through adapting various technological situations to fit a paper-and-pencil mode of testing, designing graphics to accurately illustrate various information technology processes, preparing various diagrams and charts as necessary visual references for various problem-solving and application questions, direct observation and assessment of such skills as computer operating, and various technical considerations, such as working around the limitations of a word processor in preparing camera-ready copy for assessment instruments.

Information technology studies will definitely include learning experiences with various types of computer equipment. Some of this learning will involve hands-on activities with computer terminals and peripheral equipment. Assessing students' levels of skill in using the equipment will involve teacher observations and the use of rating scales. Scoring students'

performances will require the observer (teacher) to make judgments within the scoring framework of the rating scales.

Everything considered, the task of assessing student learning in the core area of information technology will be a challenging task.

LANGUAGE ARTS ASSESSMENT

The assessment of student performance in the core area of language arts is a task which reaches beyond the rather traditional testing of reading, spelling, and language mechanics skills such as capitalization, punctuation, and word usage. The interrelated processes of thinking, reading, writing, speaking and listening together make up Utah's language arts core curriculum. These processes represent an integrated whole, parts of which are experienced in isolation, but which generally interrelate and act in concert. It can be argued that the teaching of effective language skills is a shared responsibility; these skills are taught, reinforced, and expanded repeatedly through all subject areas.

Assessment processes being developed to determine students' mastery of the language arts core curriculum include item pools and end-of-level tests. The item pools and end-of-level tests will be used to assess reading skills of students in grades 1 through 6. Tests for assessing students' writing and listening skills at grade levels 3 and 8 are also being developed.

The reading item pools and tests will include a comprehensive array of test items designed to assess phonetic analysis, including consonants and vowels; structural analysis, including word forms, word endings and word parts; vocabulary, including basic/sight/definition, contextual, functional and relational; comprehension, including literal, interpretative and

critical; and study skills, including locating information and organization information. The reading end-of-level tests and item pools are constructed in a multiple-choice test item format, with the student seeking to select the correct answer from among four alternative answers.

The writing skills tests, designed for use at grade levels 3 and 8, represent a departure from traditional testing approaches. These tests require the students to actually demonstrate their writing and listening skills in structured assessment situations. The testing procedure requires students to respond to writing "prompts," specific topics which require the students to prepare a written narrative. Also involved in the procedure are training sessions for teachers asked to score the students' writing and a set of scoring rubrics. Scoring is done using a holistic approach.

The listening skills test, designed for use in grade 3, requires the students to listen to an audio tape script and follow the tape recorded directions which they hear. There are two parts to the test and a structured answer sheet is provided for students to use.

Assessment of language arts is a formidable task, and will necessitate evaluation of student-produced written materials. In addition to the language arts core curriculum assessment tools now being developed, teachers will undoubtedly be giving spelling tests, scoring themes and research papers, evaluating students' grammatical skills and engaging in various other assessment tasks. The purpose of the language arts core is to help students attain language skills that generate and communicate thinking; assessment is a vital part of this process.

MATHEMATICS ASSESSMENT

Assessing student performance in the core area of mathematics is a task which will significantly utilize criterion-referenced testing strategies. Mathematics is a subject in the curriculum which does lend itself well to traditional measurement techniques generally used in the classroom. Test items can be in a paper-and-pencil format, centering around the functions of computation, conceptual understanding, and mathematical applications in support of the core curriculum standards.

Mathematics is a course for students which overlaps and contributes to several other academic disciplines, particularly the science and business areas. Additionally, a basic understanding of mathematics is necessary for adequate performance in most jobs and careers. The ultimate goal of successfully applying mathematical concepts and skills to various functions in daily living cannot be overemphasized by Utah educators as they design learning strategies for mastering the objectives outlined in the mathematics core curriculum.

In developing evaluation instruments for assessing a student's performance in mathematics, educators and testing specialists must take care to prepare an array of test items that will thoroughly encompass the many objectives contained in the core curriculum standards. The test items will need to deal with a range of computational skills, mathematical concepts, and situational applications. A variety of mathematical applications will be needed including the use of geometry concepts to explore the relationships of objects, the use of measurement strategies, the use of different numeration systems, the use of mathematics for daily living, and other areas of application requiring problem solving strategies as addressed by the core curriculum.

Some of the same constraints which challenge developers of assessment instruments in other curricular areas will also impact the development of test items for mathematics. These will include, though not be limited to, the following: describing mathematical situations to fit a paper-and-pencil testing mode predominantly using multiple-choice questions, designing graphics to accurately illustrate various measurement and geometric situations, preparing appropriate diagrams and charts as necessary visual references for various mathematical problem-solving questions, and dealing with technical considerations such as working around the limitations of a word processor in preparing assessment instruments for final printing.

When all is considered, the task of assessing students' mathematical performance will be demanding but also exciting and rewarding. Carefully designed test items, assessing various concepts and skills addressed by the standards and objectives of the core curriculum, can be prepared in objective, paper-and-pencil format and used to evaluate students' mastery of the mathematics curriculum. This test development work will require the knowledgeable input of teachers, evaluators and other educators skilled in mathematics.

MUSIC ASSESSMENT

The assessment of student performance in the subject area of music is indeed a formidable task. Music does not easily lend itself to such traditional measurement techniques as contained in traditional paper-and-pencil tests, both norm-referenced and criterion-referenced, which abound in the areas of reading and mathematics. Music, as an aural art, does not easily lend itself to the visual emphasis of traditional tests. Also, it can be

taught through several different approaches encompassing a broad range of objectives and styles of music. Students must be provided with opportunities to acquire a conceptual understanding of music and to acquire and apply the skills necessary to sing and play musical instruments. Conceptual understanding includes such aspects as listening, structural analysis, theory, historical and cultural development, and aesthetic appreciation. Applied musical skill speaks to creative, recreative, skill development, and performing opportunities for students. Obviously, no single measurement technique can in isolation adequately assess all these dimensions of music.

In developing item pools for assessing a student's performance in music, the developers of test items are challenged by some definite constraints. These include constructing as many items as possible of the paper-and-pencil variety, designing items which require external judgments in a manner which will maximize evaluative objectivity, focusing item content upon the literal interpretation of the objectives, and minimizing the use of audio-visual media in test items. While music is not a subject found in traditional standardized achievement tests, there are a few evaluation instruments available which have provided some guidance for the development of the core curriculum music item pool. These include: the Objectives-Referenced Test developed by the Kansas State Department of Education to assess their Comprehensive Music Curriculum; the Minnesota Educational Assessment Program for Music; and items from the music portion of the Michigan Educational Assessment Program.

The philosophy of assessment for music is that every effort must be made to evaluate a student's musical skills through direct observation of these skills and by paper-and-pencil measurements. Such tasks as having a student sight read music, perform music, demonstrate musical techniques, and listen

to music are examples of tasks which require the direct assessment of performance and will require structured, well-defined observations by an observer (teacher) to rate such performance. Scoring will require observers to make judgments within the framework of specific response options rather than simple right-wrong judgments. Other tasks, such as indicating knowledge of the historical development of music, musical instruments, musical notation, and aesthetic understanding, can be assessed through paper-and-pencil measurements. These measurements, however, will require special question stems and visual aids such as pictures of musical instruments, lines of musical notation, line drawings of the movement pattern of a conductor's baton, audio playback equipment, and various other atypical components. The task of assessing music performance is not impossible, it is simply formidable!

SCIENCE ASSESSMENT

All students, even if their future goals and occupations should be far removed from scientific fields, are now living in a world increasingly dominated by the rapid proliferation of the technological consequences of science. While the science core may be taught in a variety of ways, the basic approach is to provide students with instruction in the life, earth-space, and physical sciences in a spiral format in grades K through 6. In grades 7 and 8, these areas are covered in semester courses while in grades 9 through 12 the requirement is for one biological science course and one physical science course from among a selected group of core options which include biology, chemistry, physics, biological-earth science, physical earth science and three advanced placement opportunities.

In developing item pools for assessing a student's performance in science, educators and testors are challenged by some definite constraints. These include adapting scientific activities and situations to fit a paper-and-pencil mode of testing, designing graphics to accurately illustrate various science processes, preparing various diagrams and charts as necessary visual references for various scientific problem-solving questions and various technical considerations, such as working around the limitations of a word-processor in preparing camera-ready copies for test item pools. The creativity and resourcefulness required to successfully meet these challenges will be appreciable.

Developers of test items for the elementary and secondary science item pools must be concerned with the various concepts and processes of science and the scientific method; item construction cannot be limited to conventional science subject-matter topics alone. Science is an area of the core curriculum which will require numerous questions dealing with higher-level thinking skills as process and concept assessment cannot adequately occur by simply using information recognition and recall items. The real challenge will be to avoid essay and lengthy explanation types of questions, and prepare objective items (multiple choice, matching, free response, etc.) which will assess students' abilities to use science processes and concepts.

Critical to the science instruction given Utah students is the need to involve <u>active</u> science inquiry. The processes of science -- observing, classifying, inferring, predicting, measuring, communicating, using space/-time relations, defining operationally, formulating hypotheses, experimenting, recognizing variables, interpreting data, formulating models -- are as essential to student learning as the content and concepts. Equally essential are quality assessment processes to assess students' mastery of science.

SOCIAL STUDIES ASSESSMENT

The assessment of student performance in the area of social studies is indeed a challenging task. Social studies does not always lend itself to traditional measurement approaches commonly used in the classroom; this subject area, similar to the subjects of art and music, will require the use of visual references in order for adequate assessment to occur. One of the most commonly accepted principles of learning embodied in the social studies core curriculum is the importance of integrated learning, emphasizing the understanding of concepts and processes over the mere acquisition of isolated facts.

Stressing the mastery of integrated knowledge helps students to progress from what is known to gain some understanding of how to deal with the unknown. This integrated approach to learning also helps students to see relationships and patterns which help in making generalization, to understand the interrelatedness of the subject and skills areas, and to be a successful learner. An integrated curriculum helps students learn how to learn, and integration needs to be emphasized in the social studies fields of anthropology, economics, geography, history, political science, psychology, and sociology.

In developing evaluation instruments for assessing a student's performance in social studies, the developers of test items are faced with some constraints. There is a need to construct as many items as possible in an objective, paper-and-pencil format; design items which deal with higher-level thinking and process skills; focus item content upon literal interpretation of the objectives; prepare appropriate diagrams, maps and charts when needed as visual references for problem-solving questions; and deal effectively with technological limitations in preparing test items for final printing in item pools and end-of-level tests.

In developing evaluation instruments to assess students' performances in the social studies, care must be taken to ensure that the array of test items prepared will comprehensively address the many objectives subsumed under the core curriculum standards. Items will encompass such topics as citizenship practices and principles, basic American values, economic literacy, American governance, global awareness and geographic skills, historical events, human development, social interactions, and various other areas emphasized in social studies courses.

People are social beings who need wholesome human relationships and productive problem solving skills. Social studies helps students to fulfill these needs in a world which is becoming increasingly interdependent.

THE GREAT SEAL OF THE STATE OF UTAH
1896